

УДК 620.179.1:004

**ИСПОЛЬЗОВАНИЕ КЛАССИФИКАЦИОННЫХ
МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ
В ДЕФЕКТΟΣКОПИИ**

Б.Б. Ильясов, К.Б. Кашимов

Павлодарский государственный университет
им. С. Торайгырова, КазахстанE-mail: razorpvl@gmail.com, kashimov@gmail.com**Ильясов Борис Борисович**,
магистрант Павлодарского
государственного универси-
тета им. С. Торайгырова.E-mail: razorpvl@gmail.comОбласть научных интересов:
неразрушающий контроль.**Кашимов Кахар Бакыевич**,
инженер Павлодарского госу-
дарственного университета
им. С. Торайгырова.E-mail: kashimov@gmail.comОбласть научных интересов:
анализ данных, машинное
обучение.

В последние годы, с ростом производительности персональных компьютеров и появлением программных продуктов, позволяющих в полной её мере реализовать, применение методов машинного обучения набирает всё большую популярность. В этой статье сделан обзор наиболее популярных методов машинного обучения в контексте дефекто- и структуроскопии. Выбор методов определялся в первую очередь импакт-фактором имеющихся публикаций, а также собственным опытом применения перечисленных методов. Также упор был сделан на классификационные методы машинного обучения как имеющие наиболее низкий порог вхождения и в большинстве своём не требующие значитель-

ных знаний в области машинного обучения и его математического аппарата. Таким образом, классификация – это раздел машинного обучения, задачей которого является определение принадлежности произвольного объекта к одному из классов на основе прецедентов. В данной статье рассмотрены основные классификационные методы машинного обучения. Приведено краткое теоретическое описание, даны примеры применения классификационных методов машинного обучения в дефектоскопии.

Ключевые слова:

Дефектоскопия, машинное обучение, классификация, анализ данных, байесовские сети, скрытые модели Маркова, метод гауссовских смесей, метод k-ближайших соседей.

Введение

Поиск дефектов может быть определен как проблема классификации. Классификация может быть как жесткой (0 – нет дефекта, 1 – есть дефект) так и нечеткой (множество значений от 0 до 1). Для решения проблем классификации существует множество методов, однако самыми популярными и зарекомендовавшими себя в области дефектоскопии являются следующие – байесовские сети, скрытые модели Маркова, метод гауссовских смесей, метод k-ближайших соседей, которые подробно описаны в данной статье.

Байесовские сети

Байесовская сеть (БС) – направленный ациклический граф, где вершина соответствует атрибуту, а дуги графа указывают на наличие зависимости между этими атрибутами (рис. 1, табл. 1) [1]. Таким образом, БС – это модель для оценки совместного распределения данных.

$$p(X_1, \dots, X_n) = \prod_i (p(X_i | Pa(X_i)))$$

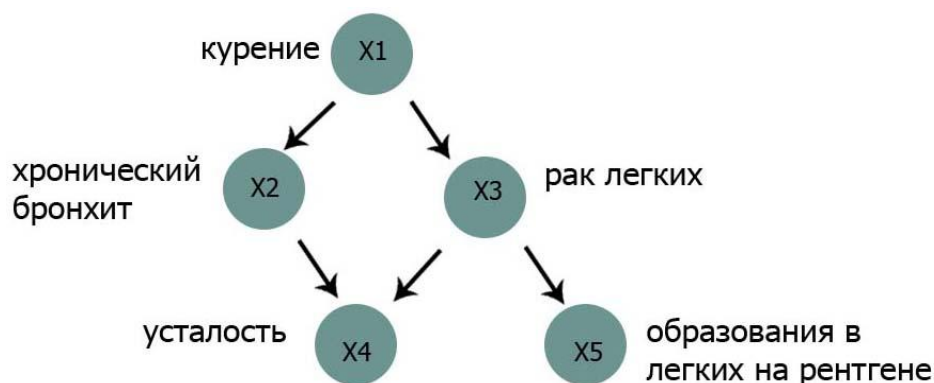


Рис. 1. Простейший пример БС

Таблица 1. Бейсовский анализ для модели, данной выше

$P(X1=no)=0.8$	$P(X1=yes)=0.2$
$P(X2=absent X1=no)=0.95$ $P(X2=absent X1=yes)=0.75$	$P(X2=present X1=no)=0.05$ $P(X2=present X1=yes)=0.25$
$P(X3=absent X1=no)=0.99995$ $P(X3=absent X1=yes)=0.997$	$P(X3=absent X1=no)=0.00005$ $P(X3=absent X1=yes)=0.003$
$P(X4=absent X2=absent, X3=absent)=0.95$ $P(X4=absent X2=absent, X3=present)=0.5$ $P(X4=absent X2=present, X3=absent)=0.9$ $P(X4=absent X2=present, X3=present)=0.25$	$P(X4=present X2=absent, X3=absent)=0.05$ $P(X4=present X2=absent, X3=present)=0.5$ $P(X4=present X2=present, X3=absent)=0.1$ $P(X4=present X2=present, X3=present)=0.75$
$P(X5=absent X3=absent)=0.98$ $P(X5=absent X3=present)=0.4$	$P(X5=present X3=absent)=0.02$ $P(X5=present X3=present)=0.6$

БС применяется в вихретоковом контроле для увеличения точности оценки геометрии дефекта [2]. Возможно использование нескольких видов классификации, таких как метод главных компонент (МГК) и нелинейный метод главных компонент (НМГК). Эксперименты показывают, что НМГК работает лучше других методов классификации [3].

С помощью метода обращения времени, основанного на БС, возможна компенсация эффекта мультимодального распространения при акустическом контроле. Результаты моделирования показывают, что возможна визуализация и определение положения дефектов с высокой точностью при использовании БС [4].

Также возможно применение гибридной системы, основанной на использовании данных с датчиков (локальные данные) и структурной информации (глобальные данные). С помощью предварительной информации, извлеченной из данных инфраструктуры, БС может помочь сделать вывод о достоверности дефекта [5].

Модели гауссовских смесей

Модели гауссовских смесей позволяют смоделировать любой набор данных (рис. 2). С увеличением количества параметров возрастает точность модели. В конечном итоге достигается компромисс между количеством параметров и точностью модели. Каждый компонент смеси

определяется так называемой ответственностью (или «весами»), которая выражается параметром π_i и параметрами гауссиан μ_i и Σ_i :

$$p(\theta) = \sum_{i=1}^K \pi_i N(\mu_i, \Sigma_i).$$

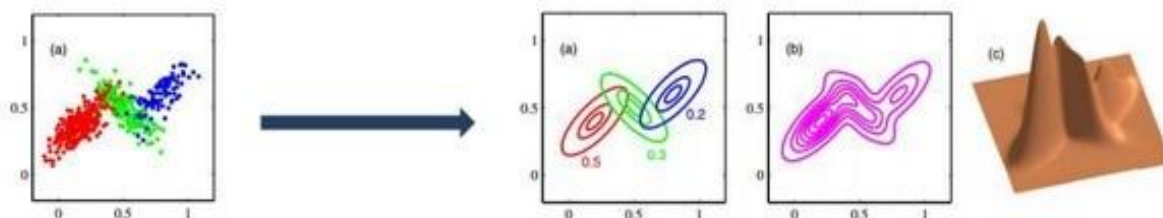


Рис. 2. Использование гауссовских смесей для создания моделей на основе данных

Оценка параметров смеси гауссиан производится с помощью ЕМ-алгоритма (англ. *expectation-maximization*).

МГС используется для сегментации изображения в системе распознавания рентгенограмм. Численные эксперименты показывают отличную производительность и точность по сравнению с методом нечеткой кластеризации k-средних [6]. МГС используется и для автоматизаций акустического контроля, а именно для обнаружения и классификации дефектов [7].

Скрытые модели Маркова

Скрытые модели Маркова (СММ) основаны на цепях Маркова, но в СММ состояния не могут наблюдаться напрямую, и каждое состояние имеет определённую вероятность эмиссии. Исходя из последовательности и вероятностей этих эмиссий, рассчитывается наиболее вероятная последовательность скрытых состояний (рис. 3).

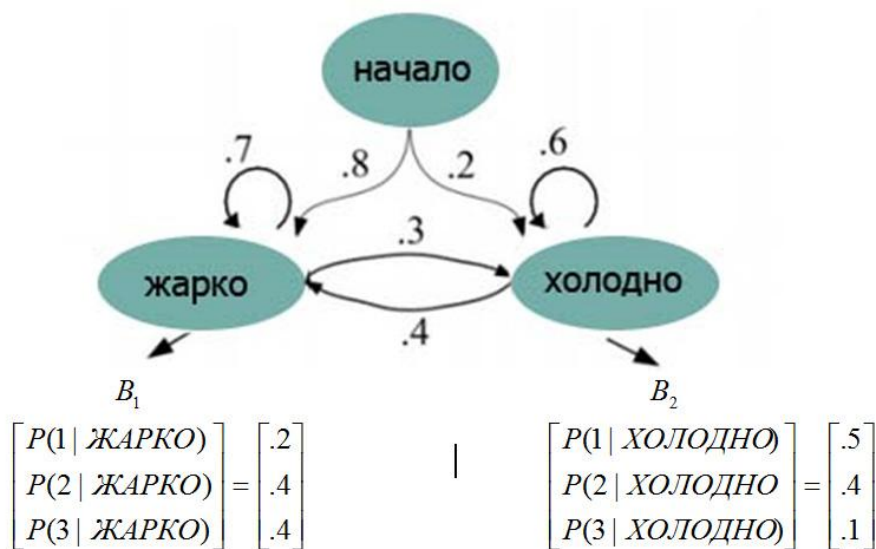


Рис. 3. СММ для сопоставления количества съеденного мороженого («эмиссии» или «наблюдения») в зависимости от погоды («скрытые состояния») [8]

СММ широко используются для распознавания изображений, что, в свою очередь, может использоваться для определения дефектов на рентгенограммах, при визуальном контроле и т. д. Данный метод позволяет снизить влияние шумов или внешних факторов, таких как уро-

вень освещения. Также СММ применяется для повышения разрешения изображений в компьютерных 3D-томограммах [9].

Такие методы, как СММ и динамическая трансформация шкалы времени, которые обычно используются в системах распознавания речи, могут быть применены и для неречевых акустических сигналов, но не являются достаточными во многих случаях [10].

Метод k-ближайших соседей

Метод k-ближайших соседей (англ. *k-nearest neighbor algorithm*, kNN, МкБС) – метод автоматической классификации объектов. Принцип МкБС основан на назначении элемента к тому классу, который является наиболее распространенным среди соседей данного элемента.

Соседи подбираются из множества объектов, классы которых уже определены, и, исходя из ключевого для данного метода значения k, высчитывается, какой класс наиболее распространен среди них (рис. 4).

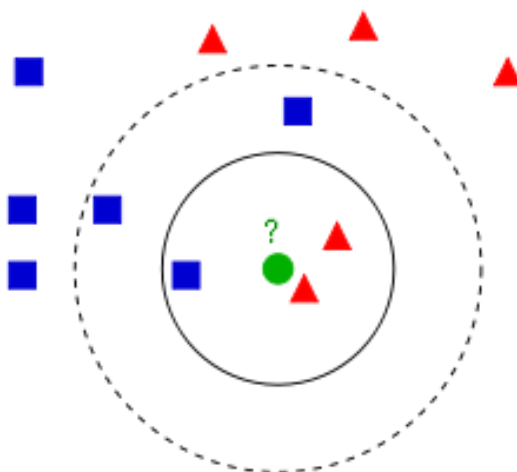


Рис. 4. Пример классификации k-ближайших соседей. Тестовый образец (зеленый круг) должен быть классифицирован как синий квадрат (класс 1) или как красный треугольник (класс 2). Если $k = 3$, то она классифицируется как 2-й класс, потому что внутри меньшего круга два треугольника и только один квадрат. Если $k = 5$, то он будет классифицирован как 1-й класс (три квадрата против двух треугольников внутри большего круга)

МкБС часто используется в системах анализа изображений. В частности, МкБС позволяет оценить пределы обнаружения дефектов в инфракрасной термографии [11]. Наряду с методом опорных векторов и вероятностной нейронной сетью МкБС является одним из наиболее эффективных классификаторов для анализа данных инфракрасной спектроскопии [12].

МкБС в вихретоковом контроле позволяет автоматически оценивать дефекты во время ручного сканирования, при этом уменьшая влияние скоростного эффекта [13].

Выводы

С развитием машинного обучения методы классификации получают более широкое распространение, в том числе и в области неразрушающего контроля. В этой статье были описаны несколько примеров наиболее удачного их применения с точки зрения авторов.

СПИСОК ЛИТЕРАТУРЫ

1. Rajesh G., Gnanasekar A., Dr. Suresh R.M. Web Service Discovery Using Semantically Annotated Belief Network // International Journal of Research in Engineering & Advanced Technology. – 2013. – V. 1.
2. Boo YE, Pingjie Huang, Mengbao Fan, Guangxin Zhang, Dibo Hou and Zekui Zhou. Geometry Parameters Estimation of Defects in Multi-layered Structures Based on Eddy Current Nondestructive

- tive Testing Technique with Bayesian Network: 17th World Conference on Nondestructive Testing. 25–28 Oct. 2008, Shanghai, China.
3. Bo Ye, Fang Zeng, Ming Li. A Bayesian Network Method for Automatic Classification of Eddy Current NDE Signals // *Applied Mechanics and Materials*. – 2013. – V. 291–294. – P. 2775–2779.
 4. Foroozan F., ShahbazPanahi S., Moallemi N., Shokralla S. Time reversal Bayesian ultrasonic array imaging for non-destructive testing // *Signals, Systems and Computers (ASILOMAR)*, 6–9 Nov. 2011, Pacific Grove, CA. – P. 1077–1080.
 5. Oukhellou L., Côme E., Bouillaut L., Aknin P. Combined use of sensor data and structural knowledge processed by Bayesian network: Application to a railway diagnosis aid scheme // *Transportation Research Part C: Emerging Technologies*. – 2008.
 6. Mekhalfa F., Nacereddine N., Goumeidane A.B. Unsupervised Algorithm for Radiographic Image Segmentation Based on the Gaussian Mixture Model // *EUROCON. The International Conference on «Computer as a Tool»*, 2007. – P. 289–293.
 7. Alireza Farhidzadeh, Ehsan Dehghan-Niri, Salvatore Salamone. Gaussian Mixture Modeling of Acoustic Emissions for Structural Health Monitoring of Reinforced Concrete Structures // *SPIE, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. – 2013. – V. 8692 (22).
 8. Daniel Jurafsky, James H. Martin. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. – 2006. – P. 174–177.
 9. Ali Mohammad-Djafari. Super-Resolution: A Short Review, A New Method Based on Hidden Markov Modeling of HR Image and Future Challenges // *The Computer Journal*. – 2009. – V. 52 (1). – P. 126–141.
 10. Tschope C., Hentschel D., Wolff M., Eichner M., Hoffmann R. Classification of non-speech acoustic signals using structure models // *IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP)*. – 2004.
 11. Hernán D. Beníteza, Humberto Loaizab, Eduardo Caicedob, Clemente Ibarra-Castanedoc, AbdelHak. Defect characterization in infrared non-destructive testing with learning machines // *NDT & E International*. – 2006. – V. 42 (7). – P. 630–643.
 12. Roman M. Balabina, Ravilya Z. Safievab, Ekaterina I. Lomakinac. Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques. *NDT & E International*. – 2004. – V. 671 (1–2). – P. 27–35.
 13. Radislav Smid, Adam Docekal, Marcel Kreidl. Automated classification of eddy current signatures during manual inspection // *NDT & E International*. – 2005. – V. 38 (6). – P. 462–470.
 14. Lauritzen S.L., Spiegelhalter D.J. Local computations with probabilities on graphical structures and their application to expert systems // *Journal Royal Statistics Society B*. – 1998. – V. 50(2). – P. 157–194.

Поступила 07.07.2014 г.